CLAIMS

1. A system for speaker modelling, said system including:

a library of acoustic data relating to a plurality of background speakers, representative of a population of interest;

a library of acoustic data relating to a plurality of reference speakers, representative of a population of interest;

a database containing training sequence(s) said training sequence(s) relating to one or more target speaker(s);

a memory for storing a background model and a speaker model for said one or more target speakers; and

at least one processor coupled to said library, database and memory, wherein said at least one processor is configured to:

- estimate a background model based on a library of acoustic data from a plurality of background speakers;
- train a set of Gaussian mixture models (GMMs) from a library of acoustic data from a plurality of reference speakers and the background model;
- estimate a prior distribution of speaker model parameters using information from the trained set of GMMs and the background model, wherein correlation information is extracted from the trained set of GMMs;
- estimate a speaker model for said one or more target speaker(s), using a GMM structure based on the maximum *a posteriori* (MAP) criterion; and
- store said background model and said speaker model in said memory.

2. The system of claim 1 wherein the MAP criterion for the speaker model is a function of the training sequence and the estimated prior distribution

3. A system for speaker modelling and verification, said system including:

a library of acoustic data relating to a plurality of background speakers;

a library of acoustic data relating to a plurality of reference speakers;

a database containing training sequences said training sequences relating to one or more target speakers;

an input for obtaining a speech sample from a speaker;

a memory for storing a background model and a speaker model for said one or more target speakers; and

at least one processor wherein said at least one processor is configured to:

- estimate a background model based on a library of acoustic data from a plurality of background speakers;
- train a set of Gaussian mixture models (GMMs) from a library of acoustic data from a plurality of reference speakers and the background model;
- estimate a prior distribution of speaker model parameters using information from the trained set of GMMs and the background model, wherein correlation information is extracted from the trained set of GMMs;
- estimate a speaker model for said one or more target speaker(s), using a GMM structure based on the maximum a posteriori (MAP) criterion, wherein the MAP criterion is a function of the training sequence and the estimated prior distribution; and
- store said background model and said speaker model in said memory.
- obtain a speech sample from a speaker;
- evaluate a similarity measure between the speech sample and the target speaker model and between the speech sample and the background model;
- verify if the speaker is a target speaker by comparing the similarity measures between the speech sample and the target speaker model and between the speech sample and the background model; and
- grant access to the speaker if the speaker is verified as one of the target speakers.

4. The system of any one of claims 1 to 3 wherein the background model directly describes elements of the prior distribution.

5. The system of any one of claims 1 to 4 wherein the background speakers and reference speakers are representative of a particular demographic selected from a population of interest including one or more of the following: persons of selected ages, genders and/or cultural backgrounds.

6. The system of any one of the preceding claims wherein the library of acoustic data used to train the set of GMMs is independent of the library used to estimate the background model.

7. The system of any one of the preceding claims wherein the extracted correlation information is stored in a library.

8. The system of claim 7 wherein the library of correlation information includes estimated covariance of mixture component means extracted from the trained set of GMMs.

9. The system of claim 8 wherein a prior covariance matrix of the mixture component means is compiled based on the library of correlation information.

10. The system of claim 9 wherein the estimate of the prior covariance of the mixture component means is determined by one or more of the following estimation methods: maximum likelihood, Bayesian inference of the correlation information using the background model covariance statistics as prior information, or reducing the off-diagonal elements.

11. The system of any one of claims 7 to 10 wherein the estimation of prior distribution of speaker model parameters is based on said library of correlation information and the background model.

12. The system of any one of claims 1 to 10 wherein the estimation of the prior distribution further includes:

a) re-training the library of reference speaker models using the estimate of the prior distribution;

b) re-estimating the prior distribution based on the retrained library of reference speaker models; and

c) repeating steps (a) and (b) until a convergence criterion is met.

13. The system of claim 3 wherein the evaluation of the similarity measure utilises an expected frame-based log-likelihood ratio technique.

14. The of system of claim 3 or claim 13 wherein the step of verification and identification furthers includes the use of post-processing techniques to mitigate speech channel effects selected from one or more of the following: feature warping, feature mean and variance normalisation, relative spectral techniques (RASTA), modulation spectrum processing and Cepstral Mean Subtraction.

15. The system of any one of claims 3, 13 or 14 wherein the speech sample from the speaker is provided to said input via a communications network.

16. The system of any one of claims 3, 13, 14 or 15 wherein the system further utilises full target and background model coupling.

17. A method of speaker modelling, said method including the steps of:

estimating a background model based on a library of acoustic data from a plurality of speakers;

training a set of Gaussian mixture models (GMMs) from constraints provided by a library of acoustic data from a plurality of speakers and the background model;

estimating a prior distribution of speaker model parameters using information from the trained set of GMMs and the background model, wherein correlation information is extracted from the trained set of GMMs;

obtaining a training sequence from at least one target speaker;

estimating a speaker model for each of the target speakers using a GMM structure based on the maximum *a posteriori* (MAP) criterion, wherein the MAP criterion is a function of the training sequence and the estimated prior distribution.

18. A method of speaker recognition, said method including the steps of:

estimating a background model based on a library of acoustic data from a plurality of background speakers;

training a set of Gaussian mixture models (GMMs) from a library of acoustic data from a plurality of reference speakers and the background model;

estimating a prior distribution of speaker model parameters using information from the trained set of GMMs and the background model, wherein correlation information is extracted from the trained set of GMMs;

obtaining a training sequence from at least one target speaker;

estimating a target speaker model for each of the target speakers using a GMM structure based on the maximum a posteriori (MAP) criterion, wherein the MAP criterion is a function of the training sequence and the estimated prior distribution;

obtaining a speech sample from a speaker;

evaluating a similarity measure between the speech sample and the target speaker model and between the speech sample and the background model; and

identifying whether the speaker is one of said target speakers by comparing the similarity measures between the speech sample and said target speaker model and between the speech sample and the background model.

19. The method of claim 17 or claim 18 wherein the background model directly describes elements of the prior distribution.

20. The method of any one of claims 17 to 19 wherein the speakers representative of a particular of a population of interest are selected from a particular demographic including one or more of the following: persons of selected ages, genders and/or cultural backgrounds.

21. The method of any one of claims 17 to 20 wherein the library of acoustic data used to train the set of GMMs is independent of the acoustic data from said speakers representative of a population of interest used to estimate the background model.

22. The method of any one of claims 17 to 21 wherein the step of extracting the correlation information includes extracting the covariance of the mixture component means from the trained set of GMMs.

23. The method of any one of claims 22 further including the step of storing the extracted correlation information in a library.

24. The method of claim 23 further including the step of estimating a prior covariance matrix of mixture component means based on the library of correlation information.

25. The method of claim 24 further including the step of estimating the prior covariance of the mixture component means is determined by one or more of the following estimation techniques: maximum likelihood, Bayesian inference of the correlation information using the background model covariance statistics as prior information, or reducing the off-diagonal elements.

26. The method of any one of claims 23 to 25 wherein the estimation of the prior distribution of speaker model parameters is based on said library of correlation information and the background model.

27. The system of any one of claims 17 to 25 wherein the step of estimating the prior distribution further includes the steps of:

a) re-training the library of acoustic data from a plurality of speakers using the estimate of the prior distribution;

b) re-estimating the prior distribution based on the retrained library of acoustic data from the plurality of speakers; and

c) repeating steps (a) and (b) until a convergence criterion is met.

28. The method of claim 18 wherein the evaluation of the similarity measure utilises an expected frame-based log-likelihood ratio technique.

29. The of method of claim 18 or claim 28 wherein the step of verification and identification furthers includes the use of post-processing techniques to mitigate speech channel effects selected from one or more of the following: feature warping, feature mean and variance normalisation, relative spectral techniques (RASTA), modulation spectrum processing and Cepstral Mean Subtraction.

30. The method of any one of claims 17 to 29 wherein the testing and training sequences are obtained via a communication network.

31. The method of any one of claims 17 to 30 wherein said target model and said background model are fully coupled.